DOCUMENT RESUME

ED 128 469 TM 005 664

AUTHOR Fink, Arlene
TITLE The Use of Data from Competency-Based Measurement: An Instructional Developer's View.
PUB DATE [Apr 76]
NOTE 14p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Achievement Tests; *Criterion Referenced Tests; *Instructional Materials; *Material Development; *Performance Based Education; Performance Tests; *Test Construction; Test Reliability; Test Validity; Validity

ABSTRACT

Attention has recently been focused on the need to design and validate instructional materials that guarantee competent student performance. In preparing such products, developers have come to rely on competency-based achievement tests. Unfortunately, expertise for constructing reliable and valid competency-based instruments is not currently available in a form than an instructional developer can readily translate into practice. The formulation of objectives to guide instrument construction is problematic for the developer because few dependable rules exist to help select and state them. Formulating items is also difficult because many developers lack training in test writing. Further, major issues arise when the developer must establish the meaningfulness of the scores that result from competency-based instruments, a difficult and costly task that requires psychometric skill and a generous budget. Finally, a discerning developer knows in advance of instrument validation that little assistance will be available from psychometricians who have long been arguing over the designs and statistics to be used validating competency-based tests. Data derived from competency-based instruments must therefore be used with caution and confirmed with additional sources of information like observations and interviews. Developers should also require test construction and psychometric expertise. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED128469

The Use of Data from Competency-Based Measurement: An Instructional Developer's View

Arlene Fink, Ph.D.
Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM005 664

A Paper presented to the National Council on Measurement in Education
San Francisco, California
April 21, 1976

ABSTRACT

Attention has recently been focused on the need to design and validate instructional materials that guarantee competent student performance. In preparing such products, developers have come to rely on competency-based achievement tests. Unfortunately, expertise for constructing reliable and valid competency-based instruments is not currently availble in a form that an instructional developer can readily translate into practice. The formulation of objectives to guide instrument construction is problematic for the developer because few dependable rules exist to help select and state them. Formulating items is also difficult because many developers lack training in test writing. Further, major issues arise when the developer must establish the meaningfulness of the scores that result from competency-based instruments, a difficult and costly task that requires psychometric skill and a generous budget. Finally, a discerning developer knows in advance of instrument validation that little arsistance will be availble from psychometricians who have long been arguing over the designs and statistics to be used when validating competency-based tests. Data derived from competency-based instruments must therefore be used with caution and confirmed with additional sources of information like observations and interviews. Developers should also acquire test construction and psychometric expertise.

In recent years, professional educators have focused attention on the creation and empirical validation of instructional materials and procedures. To them, the words "emirically validated" usually mean a guarantee that when systematically developed materials are used as intended, students will acquire important knowledge and skills and be able to competently perform educationally significant tasks. The idea that instructional products can and should ensure learner competence has been buttressed by the movement for performance-based or competency-based education and measurement, a relatively recent innovation that is thought to offer significant mechanisms for educational reform (Hood & Blackwell, 1975).

There are, unfortunately, numerous definitions of competence, and they vary greatly. Schalock and Thomas (1973) have made a useful distinction between the two most common meanings of the term. Competence, they suggest, can be equated with the mastery of knowledge and skills assumed to be necessary to perform a particular job, or it can refer to demonstrated ability to bring about outcomes specific to a given job description. There are few school situations in which instructional product developers can follow their students to their jobs to observe their performance. Thus, the most frequently used definitions of competence in instructional development are those related to the provision and measurement of knowledge and skills.

The procedures that should be used to empirically validate competency-based instructional materials have been described in the instructional product development literature (Baker, 1973; Markle, 1967). Although development experts may disagree on the specifics of instructional design, they are

united in their acceptance of the need to conduct systematic field trials of the product with a representative sample of the intended learners. They also advocate, at a minimum, the use of data from achievement tests that measure the product's effectiveness in attaining its objectives and promoting competence.

The Center for the Study of Evaluation (CSE) at UCLA is engaged in the development of numerous competency-based instructional products to train evaluators to organize and conduct evaluations. Although attempts are made through special studies to follow performance on the job, most development activities are guided by data derived from achievement tests that assess mastery of knowledge and skills. The data are used to answer questions concerning how well and how much students are learning in order to improve the product so that their competence is maximized. Thus, implicit in the design of CSE's products is an acceptance of the view that competency-based education serves its potential audience well if it provides them with the skills and knowledge presumed to be related to successful performance, and the belief that expertise exists and is available to develop reliable and valid measurements of the extent to which the skill and knowledge are acquired. It is the latter belief--that expertise is available for developing reliable and valid competency-based measurement devices like achievement tests--which this paper explores from the point of view of one of CSE's instructional developers. To do this, competency-based measurement in instructional product development will be defined, and issues in the development and validation of competency-based achievement tests will be discussed.

Competency-Based Achievement Tests and Instructional Product Development

Competency-based achievement tests are designed to provide a measure of the extent to which an instructional product's objectives have been achieved in equipping learners with specific competencies.[1] They have three essential characteristics:

1. they are based on clearly-defined performance objectives representing competencies;

2. test items are specifically designed to measure the objectives;

3. scores are interpreted in terms of attainment of a pre-set criterion or standard of performance.

Formulating Objectives/Competencies

One of the instructional product developer's basic responsibilities is to formulate the performance objectives that frame the product and guide achievement test development. In competency-based education, this is usually done by obtaining the advice of experts, analyzing the subject area to be learned, conducting a needs assessment, and by monitoring student progress during field tests of the product. Several problems invariably arise when formulating objectives. The most common include determining which ones to select; whether to state them in specific or general terms; how to sequence them (from simple to complex? from abstract to concrete?); and finally, how to interpret field test data in terms of the importance and comprehensiveness of objectives.

The latter problem, interpreting field test data, has become a particular

ly irksome but integral part of the life of any developer concerned with competency-based measurement. For instance, consider a developer faced with a situation in which all students correctly answer all questions on a competency-based pretest. In theory, this would suggest that the pretest was too easy and needed revision or that the objectives were irrelevant since they had already been achieved. But wait, and consider another developer confronted with a situation in which all students do poorly on the competency-based posttest. In theory, this would suggest that the materials used were in need of revision since they were not teaching successfully, that the posttests were too hard, or that the objectives were inappropriate for the learners who were incapable of achieving them.

Unfortunately for both developers, it is often impossible to decide which interpretation is correct, that is, whether the test is at fault in being too easy or too hard or whether the materials need revision. This ambiguity exists because the measures used to test the efficiency of the instructional product typically have not been validated themselves. In these situations the two developers' data are truly uninterpretable, but what if they are all the developer has? There are very few instances in which developers have access to validated instruments that yield credible information. In fact, in most cases, product development and test development occur together, with data from one rough draft used in an intuitive way to refine the other, a circumstance closely resembling the inauspicious one in which the blind lead the blind.

Given the problems in formulating objectives and in interpreting informa-

tion about them, not to mention the costs and effort involved in field tests, what should the concerned developer do? Unfortunately, at present, few rules are available although attempts have been made to describe development activities and to provide general guidelines for them (Popham & Baker, 1971).

Formulating Items for Competency-Based Measures

Once objectives have been formulated, the developer is then confronted with the task of constructing items to measure achievement. This task is a very difficult one for developers who are frequently individuals with little formal training in test writing and validation, and who may be skillful instructors, but poor item writers and analysts. Further, even having knowledge of psychometrics does not vitiate many unanswered questions about competency-based measurement that seem to exist to complicate the developer's job. For example, how many items are needed to measure an objective? Even highly specific objectives have potential item pools of well over several thousand items (Bormuth, 1970; Hively, 1970; et al, 1973). Should the developer sample items? If so, how? A related question concerns the selection of items according to their difficulty. Thus, the developer must not only decide how many items are needed to measure an objective, but also determine the level of difficulty for each item. Given the problems involved in test generation, the developer would be well advised to call in experts or to obtain "item forms" or "shells" to guide test development (Cronbach, 1971; Hively et al, 1973; Skager, 1973). However, experts are expensive and currently available item-writing rules are far from generally available or even applicable to

many competency-based instructional development concerns.

Score Interpretation

Competency-based measurement provides descriptions of what individuals know or are likely to be able to do and reports scores in terms of performance standards. A major problem often arises, however, in establishing the meaningfulness of the standards since to be legitimate and credible, they should preferably be empirically justifiable. For example, a score of 7 out of 10 items will only have meaning if systematic study has shown that those who receive the score can actually do something that others who have not reached this level cannot do. Needless to say, such systematic investigation is costly and requires ability to conduct testing research. Few instructional product developers have had the opportunity to participate in this kind of research, and few development budgets include provisions for it. In its absence, many developers must continue to unhappily rely on arbritrary scores like 7 out of 10 to estimate gains in competence for information about their product.

Establishing the Technical Excellence of Competency-Based Measurement

It is axiomatic that all tests and measures must be field tested before decisions can be based on them. Thus, when the construction of an achievement test's items is complete, the test must be validated. The purpose of the validation is to ensure the quality of each item, and the test as a whole, and to be able to declare the test to be a valid and a reliable instrument. However, developers are usually not trained in the vicissitudes of conducting and costing test validation. To further

complicate an already difficult task, debate continues over the appropriateness of classical indexes of reliability and validity, which are based on variance, for instruments that are used in situations that produce little variation in scores (Popham & Husek, 1969). Such situations are well known to instructional product developers who will undoubtedly be familiar with the following occurrence: Before instruction, none of the students are competent with respect to the objectives and they all receive low scores on the pretest, whereas after instruction, they all receive high scores. Most theoreticians would argue that this is the individual or idiosyncratic case, that variation is inevitable in most instructional situations, and that therefore, classical indexes are appropriate (Harris, 1973). Nonetheless, as interesting and important as the issue may be, developers cannot be expected to take sides in the debate and are to be forgiven if they regard data from competency-based measurement with a degree of confusion and distrust.

Use of Competency-based Measurements: Conclusions

At present, expertise for developing reliable and valid competency-based instruments is not available in a form that an instructional developer can readily translate into practice. The formulation of objectives, construction of items, interpretation of scores, and determination of reliability and validity remain problematical. Thus, data obtained from competency-based measurement, which are to be used to improve products and describe performance, must be viewed with extreme care, if not actual suspicion. The developer should take care to confirm any findings from

competency-based measurement by relying upon additional sources of information like students' comments and observers' ratings. Further, developers should become acutely aware of the part instrument development is likely to play in product design, and be ready to acquire psychometric expertise to prepare them for the problems that are likely to emerge. In the meantime, data derived from competency-based measurement should be handled with care and at the developer's own risk.

Footnote

[1]Competency-based achievement tests and criterion-referenced tests are identical in most respects. The major difference between them is that competency-based tests are always directly job-related, while criterion-referenced tests can also be used to measure academic skills that may be only indirectly related to job performance. Thus, competency-based tests are really a type of criterion-referenced testing.

References

Baker, E. L. The technology of instruction. In R. Travers (Ed.) Second handbook of research on teaching. New York: Rand McNally, 1973.

Bormuth, J. P. On the theory of achievement test items. Chicago: University of Chicago Press, 1970.

Cronbach, L. J. Test validation. In L. Thorndike (Ed.), Educational measurement (2nd ed). Washington, D.C.: American Council of Education, 1971.

Harris, C. Comments on problems of objective based measurement. Paper presented at Annual AERA meeting, New Orleans, 1973.

Hively, W. Introduction to domain referenced achievement testing. Symposium presentation, AERA, Minnesota, 1970.

Hively, W., Maxwell, G., Rabehl, G., Sension, D., & Lundin, S. Domain referenced curriculum evaluation: A technical handbook and case study from the MINNEMAST project. CSE Monograph Series in Evaluation, Volume 1. Center for the Study of Evaluation. University of California 1973.

Hood, P. & Blackwell, L. An assessment system for competency-based education: The educational development, dissemination, and evaluation training program. San Fransico, Far West Laboratory, Dec., 1975.

Markle, S. M. Emipirical testing of programs. Programmed Instruction. The sixth yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1967.

Popham, W. J. & Baker, E. L. Rules for the development of instructional products. In Robert L. Baker and Richard L. Schultz (Eds) Instructional product development. Beverly Hills, California: Van Nostrand & Reinhold Co.; 1971.

Popham, W. & Husek, T. R. Implications of criterion referenced measurement. Journal of Educational Measurement, 1969, 6(1), 1-9.

Schalock, H. D. & Thomas, G. P. Some observations on the contribution of the Oregon studies in Educational RDD & E to the development of competency-based personnel development programs. Monmouth, Oregon: Teaching Research, 1973.